

Identifying the Major Traits of Ethnic Clustering in England and Wales from the 2011 Census

Guy Lansley¹, Yiran Wei¹ and Tim Rains²

¹*Department of Geography, University College London, London, UK*

²*J Sainsbury's plc, Coventry, UK*

Introduction

Ethnicity has long been a major subject in the realm of social research in the UK. It describes an umbrella of characteristics that are based on the premise that groups of people who have their roots in common ancestry, religion, nationality, language and territory share similar traits and culture (Bulmer, 1996). The definition, measurement and classification of ethnicity has attracted on-going debate in amongst researchers due to its multidimensional, subjective and complex nature (Mateos *et al.*, 2009).

The 2011 Census for England and Wales identified that the population is becoming ethnically more diverse, largely due to immigration and higher fertility rates amongst most ethnic minority groups compared with the national average (Simpson, 2013). Typically minority groups residentially cluster within urban areas due to a range of structural social and economic forces (Finney, 2013). While minority ethnic groups are now dispersing (Stillwell and Hussain, 2010), many metropolitan neighbourhoods across the country are still commonly associated with particular ethnicities.

The spatial segregation of ethnic minorities within urban areas in Britain and its effects on wider society have been a major focus of debates in both politics and the media, and a topic of considerable academic interest (Peach, 1996). Despite this, a single indicator of neighbourhood ethnic composition has not been produced at a small area level within England and Wales. As the diversity of the population increases, it would be beneficial to find means to easily identify the local composition of ethnic and cultural groups in order to improve local service provision. This could include, for example, improvements to local shopping facilities, in particular, grocery store provision. Many minority ethnic and cultural groups in Britain have distinctive food consumption habits which emanate from their cultural origins (Uskul and Platt, 2014). Therefore, understanding a basic segmentation of ethnic composition at a small area level across England and Wales would be useful to supermarket planners aiming to make their stores more relevant to the shopping requirements of their surrounding catchments.

Using data at the output area (OA) level from the 2011 Census, this research aims to identify major spatial variations in ethnic composition between neighbourhoods across England and Wales. This has been achieved by the creation of a Cultural, Ethnic and Linguistic Output Area Classification (CELOAC), a composite indicator which comprises a range of variables that describe cultural heritage such as ethnicity, religion, migration and language. The classification has then compared with the total sales of a selection of ethnic origin foods using supermarket customer loyalty data also recorded at the OA level to identify the association between ethnic composition and food consumption.

What is ethnicity?

Ethnicity can be an intangible concept. Definitions range from primordialist theories, which describe ethnicity as a physical outcome derived from ancestry, to constructivist theories, which perceive ethnicity as a social construction (Wan and Vanderwerf, 2009). Following his study of the first question of ethnicity used in a UK Census, Bulmer defined an ethnic group as a "collectivity within a larger population having real or putative common ancestry, memories of a shared past, and a cultural focus upon one or more symbolic elements which define the group's identity"(1996:35). Such elements included shared

kinship, religion, language, location, nationality, physical similarities from ancestry (Bulmer, 1996). These attributes, individually, may not always pose as a useful indicator of ethnicity, failing to acknowledge its multidimensionality. Researchers from different fields, but most notably those investigating ethnic inequalities, have agreed that using a range of attributes to identify ethnicity is far more appropriate than considering just one basic measurement (Bhopal 2004; Gerrish 2000; McAuley *et al.* 1996; Mateos, 2014b).

Large scale and historical migration flows have confused traditional conceptions of ethnic groups. No longer can ethnicity be defined or identified by a common geography (Levinson 1998). Even self-defined ethnicities can be unstable. One study found that 4% of persons recorded a different ethnicity in the 2011 Census in England and Wales, compared with how they recorded themselves in 2001. The rate of instability within the Irish ethnic group was as high as 26% (Simpson *et al.*, 2014). Traditional definitions of ethnicity are therefore not a robust indicator of cultural identity.

Ethnicity and residential segregation

The basic premise of geodemographics is that 'birds of a feather will flock together' (Flowerdew and Leventhal, 1998). This statement applies to the multitude of geodemographic facets which together describe a community, notably including ethnicity. Consequently, it is not surprising that ethnicities are not evenly distributed across the country. Peach (2006) identified two main theories of ethnic minority residential distributions; multiculturalism and assimilation. Multiculturalism refers to the preservation of segregated neighbourhoods, often despite economic assimilation, due to cultural ties and other social forces. Assimilation refers to the gradual absorption of minorities into mainstream society.

Currently, within England and Wales' urban areas, there are several areas which can be considered ethnically segregated despite a general trend towards assimilation evident amongst the majority of ethnic communities (Stillwell and Hussain, 2010). This was an issue popularised by Trevor Phillip's (then chair of the former Commission for Racial Equality) who spoke out in 2005 about his fears that Britain was 'sleepwalking into segregation' (Finney and Simpson, 2009).

There are many reasons why ethnic minorities often residentially cluster, and in extreme cases, form ethnic enclaves where neighbourhoods become culturally distinctive from mainstream society (Portes and Jensen 1987). Typically, migrants settle in inner city areas and over generations, develop into segregated ethnic minority communities. Inner city locations often fulfil the desire to reside near employment, usually available in city centres and they also often provide cheap, high density housing (Vaughan, 2007). Johnston *et al.* (2007) researched segregation in five western Anglophone countries and argued it was a consequence of three main processes: disadvantage, discrimination and choice. Members of ethnic minority groups are more likely to be disadvantaged, in terms of access to employment, education and skills, and hence well paid jobs and housing (Johnston *et al.*, 2007). In some cases, these disadvantages can lead to social exclusion and prevent ethnic minority groups from participating in mainstream society. The capacity for these disadvantaged and excluded groups to relocate into the wider urban area therefore, is greatly restricted.

Social networks often provide immigrants with social capital that can be transferred to other tangible forms. It is beneficial therefore, to retain such social links (Abrahamson, 1995; Douglas 1990). Traditionally, generations of immigrants have followed their predecessors to locations in which they can benefit from social and family networks, frequently in terms of feelings of security, and economic and housing opportunities (Massey, 1990). This is especially important where cultural differences may restrict such opportunities elsewhere (Vaughan, 2007). Over generations, these factors can reinforce and develop ethnic identity. Similarly, Simpson and Finney (2009) reviewed the concept that people stay close to where there is plenty of social support and that this in turn, reinforces the grouping of ethnic minority communities. Those from the same ethnic or cultural backgrounds tend to be more likely to be socially supportive to one another

(Simpson, 2004). For example, participation in religious and other group-related activities provide incentives to cluster for some minority groups, as demonstrated by the Jewish community in London. Most members would prefer to live near their cultural institutions and businesses such as synagogues, kosher butchers and Sunday schools (Vaughan, 1997).

Data

Data for the classification was obtained from the 2011 Census for England and Wales at the output area level. Output areas (OAs) are the smallest geographical unit for which data is available from the 2011 Census. There are some 180,000 OAs across England and Wales with an average population of 309. Data from the Scottish and Northern Irish 2011 Censuses were not analysed owing to their variables not being standardised with the English and Welsh releases, and they did not release such a granular list of responses pertaining to cultural identity. The Census is the most valuable source for information on cultural and ethnic compositions at a small area level (Finney, 2013). As with the 2001 Census in England and Wales, the 2011 survey produced data tables on country of birth, ethnicity and religion. In addition, the 2011 Census also contributed new data on main language, national identity, and year of arrival in the UK. Each of these new datasets can contribute to better understanding of cultural identity (Mateos 2014a).

A total 435 variables relevant to CELOAC are available from the 2011 Census for England and Wales at the output area level. They cover 7 determinants or dimensions of cultural identity including country of birth, ethnic group, religion, main language, proficiency in English, age of arrival in UK, and length of stay in UK (Table 1).

| Census Table | Name |
|--------------|-------------------------------|
| QS203EW | Country of birth (detailed) |
| QS204EW | Main language (detailed) |
| QS205EW | Proficiency in English |
| QS208EW | Religion |
| QS211EW | Ethnic group (detailed) |
| QS802EW | Age of arrival in the UK |
| QS803EW | Length of residence in the UK |

Table 1. 2011 Census tables selected for CELOAC.

The first category of census data describing cultural and ethnic identity analysed by the study was Ethnic group. In the 2011 Census survey this was collected in the form of written responses which were subsequently classified by the Office for National Statistics (ONS) and disseminated as a classification of 250 individual ethnic groups. This variable is an imperative indicator of cultural identity as it signifies the ethnicity each individual identified with in the 2011 Census. However, as a record of self-defined ethnicity, it can be considered to be to some extent, an unstable measure. Therefore, additional variables were used to identify ethnic identity. The second set of variables included in the analysis was country of birth. This records first generation migrants' origins and it is an important foundation of cultural identity. Unfortunately, the Census did not publish information on the family origins of second and third generation migrants, these individuals were simply recorded as British born. Two variables sets on language were also included, one referring to English language proficiency and the other to main language. Main language is a good proxy for cultural identity and integration amongst migrant communities. Main language is not always constricted by national borders and may span several countries whilst other languages may be isolated to distinctive regions within nations. Over 4.15 million persons in England and Wales (7.7% of the population) did not record English as their main language. English language proficiency is an important indicator as insufficient English communication skills can act as a barrier to cultural integration with the wider society. Religion is also an important aspect of cultural identity. Amongst certain cultures, religion where it might be a crucial foundation of social networks and communities which share distinctive norms and behavioral patterns. Finally, the study also considered variables on

the length of stay in the UK (of first generation migrants only) and their age of arrival. This could be useful as immigrants are generally more likely to identify with a host culture the longer their residency, particularly those who migrated at a young age (Cheung *et al.*, 2011). Nevertheless, cultural absorption is also a consequence of individual experiences of integration and their exposure to mainstream society.

Many of the individual variables from the seven census tables represented very small populations. Variables with total populations below 10,000 were aggregated into broader groups based on their global regions of origin or removed altogether if they were considered too distinctive to merge. Smaller populations could skew the results later in the methodology and ultimately they are only applicable to a tiny proportion of the population (Vickers and Rees, 2007). Following this step, only 134 individual variables remained.

Methods

The methodological approach for this study draws heavily from the existing literature surrounding conventional geodemographics (Harris *et al.*, 2005). And most notably, the open source Output Area Classifications (OAC) produced by the University of Leeds (2001 Census edition) and University College London (2011 Census edition) in conjunction with the ONS (Gale *et al.*, 2015).

Like both Output Area Classifications, CELOAC was built using a k-means clustering of multivariate Census data at the output area level. Prior to running the clustering, the data needed to be standardised to give each variable an equal weighting and to ease data interpretation. Following this, tests to ensure the variables were appropriate for the classification and were not unjustifiably skewing the results were pursued. Our methodological steps are outlined in figure 1.

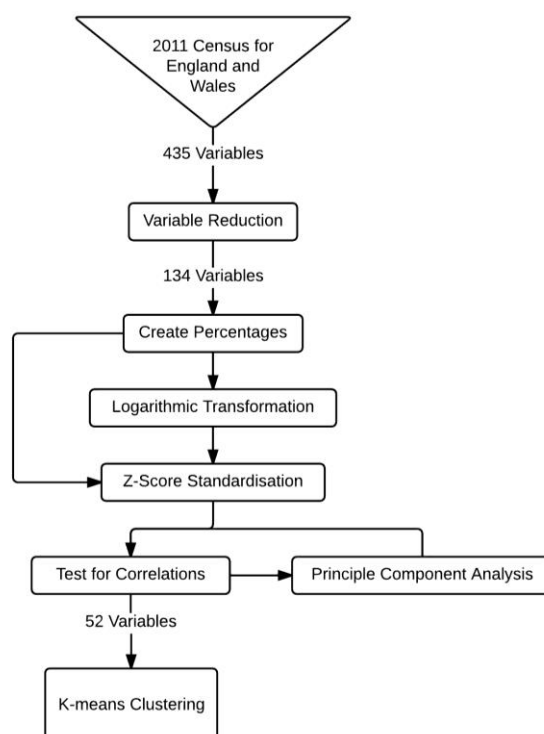


Figure 1. The methodological steps taken to produce CELOAC.

The variables initially needed to be standardised in order to reduce the effects of outliers on the univariate distributions of each variable (Milligan, 1996). Many of the individual variables were positively skewed, largely due to low counts and a tendency for cultural groups to cluster (Finney and Simpson, 2009). Therefore, natural log transformations for these cases were implemented so that the data was transformed to become roughly

symmetric and near normal. In addition, Z-score standardisation was considered so that each variable was presented on a common scale of standard deviations from mean.

Two steps were taken to gauge the appropriateness of the remaining variables. A Pearson's correlation matrix for the dataset was created to identify any variable pairs which may share a high association. The inclusion of pairs of variables with strong correlations within a dataset is undesirable for cluster analysis because they represent data redundancy and may give the same phenomenon a higher weighting (Vickers and Rees, 2007). A Pearson's correlation coefficient (r) is an indication of the direction and a measure of the strength of the association between the two variables. For this paper, any two variables with coefficients greater than ± 0.8 were considered to be highly correlated. Of pairs of variables which correlated highly, either the smallest was removed from the variable selection or they were merged into 'other' groups if both variables were from the same census table and represented similar cultural groups.

To make the final model more parsimonious, a Principal Component Analysis (PCA) was implemented to measure the influence of each of the variables across the whole sample. PCA can be used to aid variable reduction without disturbing its main features, it can also be used to identify erratic variables (Rencher, 1996). Whilst the principle components produced by the PCA were not used in the classification as they would create issues with the later data interpretation, the model was instead used to inspect the data. The model tells us the degree to which each variable can be associated with the underlying principal components (Rummel, 1970). By producing a component loading matrix and a communality coefficients table unsuitable variables could be identified and then removed so variable redundancy was reduced from the final model (Meyers *et al.*, 2006).

In total 52 variables were selected for the classification (table 2). The variable with the smallest population out of the final selection, Russian language, represented over 67,000 persons.

| 2011 Census Table | No of original variables | No of aggregated variables | No of final variables |
|-------------------------------|--------------------------|----------------------------|-----------------------|
| Country of birth | 57 | 49 | 15 |
| Ethnic group | 250 | 40 | 18 |
| Main language | 92 | 20 | 7 |
| Proficiency in English | 5 | 5 | 1 |
| Religion | 9 | 8 | 7 |
| Age of arrival in the UK | 17 | 7 | 2 |
| Length of residence in the UK | 5 | 5 | 2 |
| Total | 435 | 134 | 52 |

Table 2. The number of variables from each census variable table used to produce CELOAC at different stages of the methodology.

Clustering method

The final 52 variables were then merged into a single composite measure using a K-means clustering algorithm. Statistical clustering constructs groups of the most similar cases based on the overall similarities and dissimilarities as conveyed through the variables. K-means is most commonly used in geodemographics. It is a top down approach whereby the number of cluster groups is predefined. K-means is an iterative relocation algorithm based on an error sum of squares measure (Harris *et al.*, 2005). The equation is listed below:

$$SSE = \sum_{j=1}^k \sum_{i=1}^n \|x_i^{(j)} - c_j\|^2$$

The algorithm seeks to reduce the sum distance between each data point $x_i^{(j)}$ and their respective cluster centre c_j . Figure 2 illustrates the basic algorithm process of k-means clustering. It starts by randomly allocating seeds across a multidimensional space as defined by the variables, each case is then assigned to the nearest seed centroid to create a cluster. The centroid is then moved to the mean location of all of the cases within its current cluster. Each case is then re-assigned to clusters based on the distance to the nearest of the new centroid locations. This process repeats iteratively until the centroid locations cannot be moved as an optimum solution has been reached (Harris *et al.*, 2005).

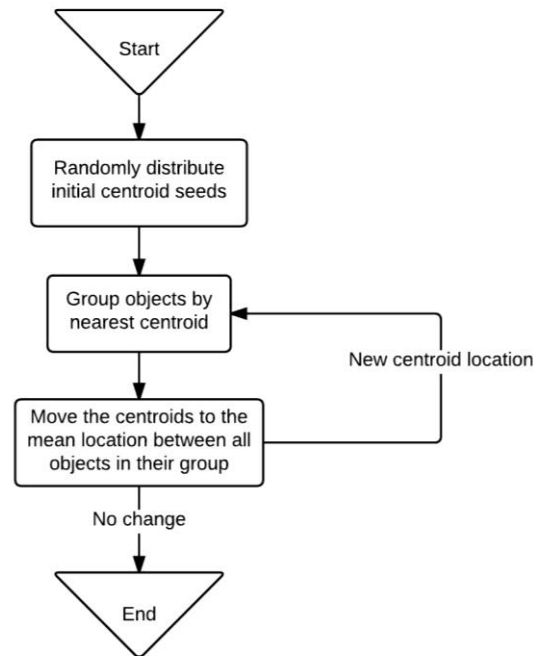


Figure 2. The process of the k-means algorithm.

The number of cluster groups to be produced had to be determined by the researchers. Different numbers of groups can create very different results. The principles that were used to choose the number of cluster groups for this classification were similar to those used by Vickers and Rees (2006). The aims were to produce clusters which were well representative of all OAs within them, but, at the same time, as distinctive as possible from all other groups. Of course, the higher the number of groups the higher the likelihood of creating groups which are truer representations. However, this also makes the model harder to interpret, and often groups can be difficult to distinguish. To put it in perspective, the 2011 OAC has 8 supergroups (which contain a hierarchy of groups and subgroups), whilst the current ACORN classification produced by CACI consists of 5 groups at its top level. Two measures of the cluster distributions from different k solutions have been presented. First, the average distance to the cluster centre. While the more clusters produced reduced the average distance across the whole sample. The second measure looked at the overall variation in the sizes of the clusters in terms of the number of OAs they represent. From observing these distributions it was decided to pursue an eight cluster solution.

Results

The CELOAC consists of 8 culturally distinctive groups. Two groups combined comprise just over 70% of OAs in England and Wales. Both contain higher proportions of the White British ethnic group than the remaining population, with rates of 88.5% and 96.2% respectively. As the focus of this research is on foreign origin ethnic groups the two white British clusters have been merged for the remainder of this paper (group G).

From looking at the cluster centres for each group (expressed as z-scores relative to the overall average), a good understanding of the cultural composition of each group can be achieved (figure 3). The classification produced 6 cluster groups distinguished by a higher presence of ethnic minorities, hereafter be labeled the minority clusters, and one larger group consisting of a homogenously White British population. The names of each group correspond with the most common cultural and ethnic group(s) based on the mean z-scores. They are only intended as labels to aid interpretation in this paper and they should not be considered derivative of each inhabitant.

Group A (Pakistani & Bangladeshi) is dominated by South Asian ethnic groups including Pakistani and Bangladeshi ethnicities. It also has the greatest concentration of those who identified themselves as Muslims. Group B (India & South Asia mix) has a heavy concentration of those of Indian ethnicity, and also those of other South Asian countries. It has the lowest percentage of white British ethnicity of all the clusters and likewise it shares the highest proportion of those who cannot speak English. Group C (Black African and Caribbean) is clearly characterised by an overrepresentation of Black African and Caribbean ethnic groups. Group D (Non-British White) has high proportions of those from European or other Anglophone nations. Group E (Middle Eastern & East Asian) has high proportions of those from Arabic and East Asian nations, many of which are affluent countries of origin. There is also a relatively high rate of those from other developed nations around the world. Group F (Mixed Ethnic Group) includes a more diverse range of ethnicities. It is the most assimilated of the minority groups but the White British population still represents over 70% of the population in these neighbourhoods. Finally, Group G (White British) is most commonly represented by homogenous White British communities.

| Group | Number of OAs | Percent |
|--------------------------------|----------------------|----------------|
| A: Pakistani & Bangladeshi | 8168 | 4.50 |
| B: Indian & South Asian mix | 4547 | 2.51 |
| C: Black African & Caribbean | 8068 | 4.45 |
| D: Non-British White | 5476 | 3.02 |
| E: Middle Eastern & East Asian | 4277 | 2.36 |
| F: Mixed | 20610 | 11.36 |
| G: White British | 130262 | 71.81 |

Table 3. The number of OAs in each CELOAC group.

From looking at the size of each of the clusters, the most notable distinction is that the Group G (White British) represents over 70% of OAs in England and Wales (table 3). Although advocates of geodemographic classifications would identify such a size disparity as unfavourable (Harris *et al.*, 2005), the methodological approach was robust. Instead, what it identifies is that less than 30% of OAs are culturally distinctive from the rest of the UK, which are largely characterized by more homogenously White British neighbourhoods. This result is reasonable as the White British population is known to comprise 80% of the total population, and there is a disassociation between this group and minority ethnic groups at the neighbourhood level (Finney and Simpson, 2009).

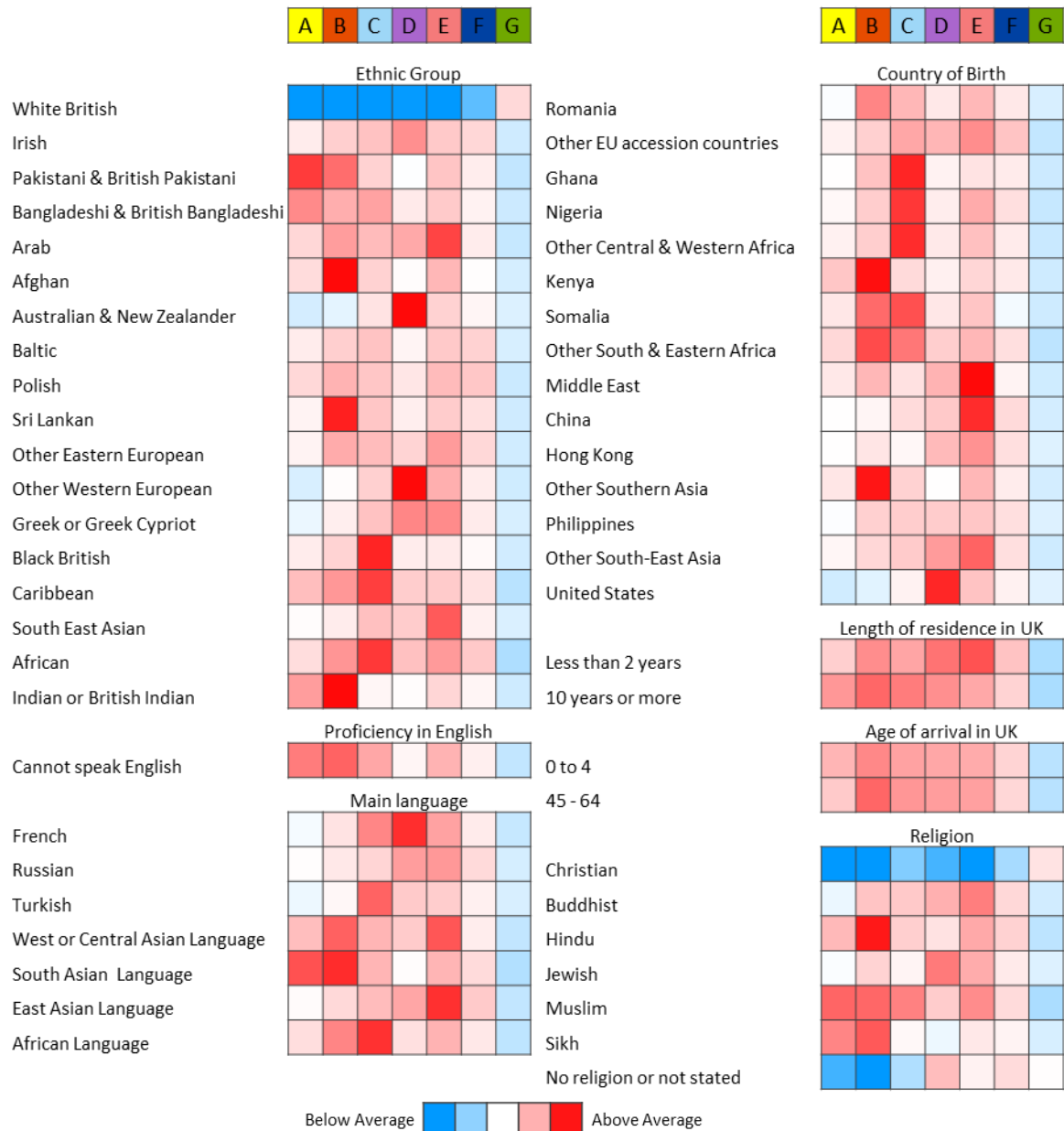


Figure 3. Cluster centre results for the 7 CELOAC groups. The colours indicate the direction and magnitude of each variable within the groups.

As the Z-scores do not convey the actual proportion of groups relative to the rest of the local population, the total percentages of large ethnic groups from the 2011 census within each of the CELOAC groups have by displayed in table 4. Despite the White British ethnic group representing over 80% of the population, they are a minority in four of the groups. The table also suggests that ethnic minorities are more likely to settle in neighbourhoods with other minority ethnic groups, rather than within White British communities.

| Ethnic Group | CELOAC Group | | | | | | | England and Wales |
|-------------------|-------------------------|--------------------------|---------------------------|-------------------|-----------------------------|-------|---------------|-------------------|
| | Pakistani & Bangladeshi | Indian & South Asian mix | Black African & Caribbean | Non-British White | Middle Eastern & East Asian | Mixed | White British | |
| White British | 43.29 | 22.50 | 33.14 | 53.73 | 42.93 | 72.15 | 93.01 | 80.49 |
| White Irish | 1.21 | 1.70 | 1.80 | 2.82 | 1.71 | 1.47 | 0.64 | 0.95 |
| Other White | 4.69 | 8.62 | 12.40 | 20.63 | 14.24 | 7.96 | 2.06 | 4.37 |
| Mixed & multiple | 3.43 | 3.81 | 6.64 | 4.88 | 4.53 | 3.49 | 1.30 | 3.12 |
| Indian | 9.84 | 25.14 | 3.01 | 2.53 | 5.55 | 3.06 | 0.73 | 2.52 |
| Pakistani | 20.23 | 9.36 | 2.85 | 0.78 | 3.46 | 1.42 | 0.34 | 2.01 |
| Bangladeshi | 5.87 | 2.83 | 3.10 | 1.10 | 1.96 | 0.66 | 0.14 | 0.80 |
| Chinese | 0.66 | 0.90 | 1.35 | 2.09 | 5.89 | 1.30 | 0.30 | 0.70 |
| Other Asian | 3.11 | 10.24 | 3.92 | 2.88 | 5.66 | 2.49 | 0.43 | 1.49 |
| Black ethnicities | 5.49 | 10.81 | 27.80 | 5.30 | 7.72 | 4.60 | 0.70 | 3.33 |
| Arab | 0.84 | 1.59 | 1.14 | 1.35 | 3.67 | 0.49 | 0.10 | 0.41 |
| Other | 1.25 | 2.38 | 2.69 | 1.85 | 2.60 | 0.77 | 0.17 | 0.59 |

Table 4. The actual percentage of ethnic groups within each CELOAC group, and England and Wales.

The geographic distribution of CELOAC groups

Mapping the distribution of the CELOAC groups in England and Wales reveals differences between the geographies of minority groups and the White British group (Group G)(figure 4). Expectedly, the minority CELOAC groups are largely concentrated in urban areas, particularly inner cities, whilst, Group G encompasses the vast majority of rural England and Wales, and many suburban areas.

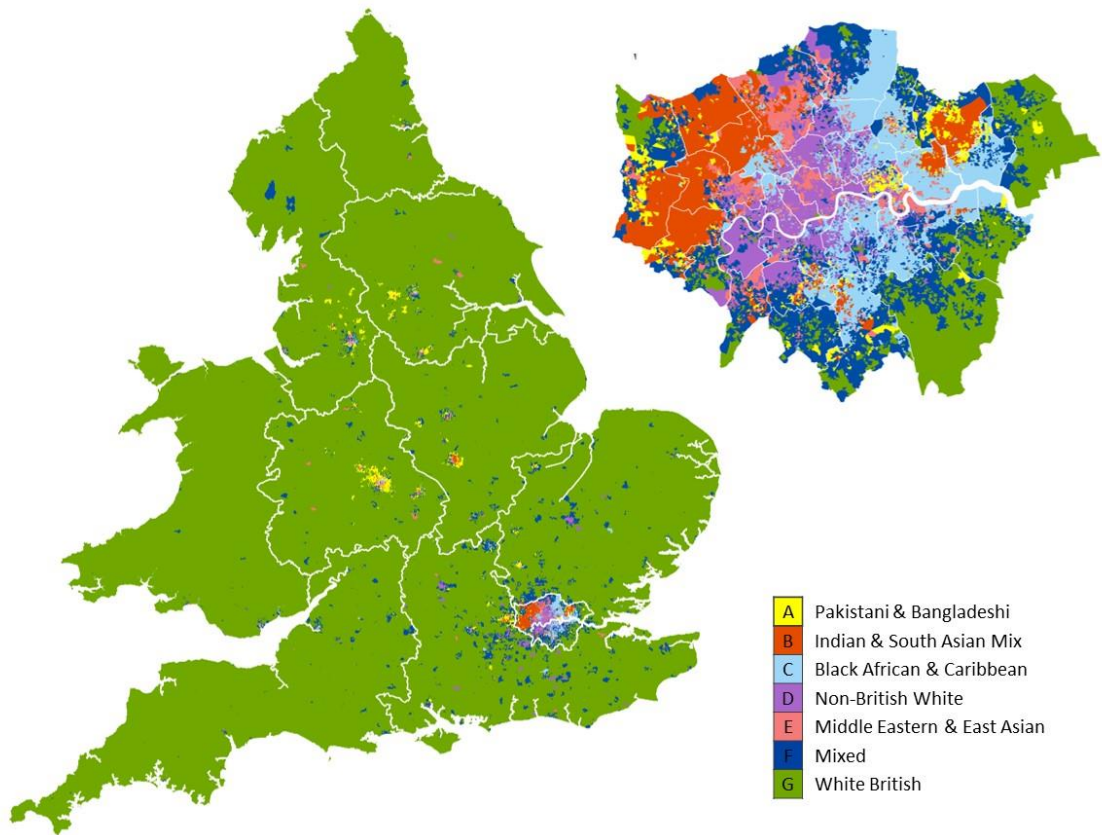


Figure 4. 2011 Cultural, Ethnic and Linguistic Output Area Classification for England and Wales

London is visibly the largest nuclei for the minority groups. There are also concentrations in other large cities which are known to have attracted large proportions of international migrants such as Birmingham, Leicester and Leeds (Dustmann *et al.*, 2011).

Regional variations

There are also distinctive regional variations in CELOAC groups across England and Wales (table 5).

| | Region | | | | | | | | | | England and Wales |
|---------|------------|------------|--------------------|---------------|--------------|-------|------------|------------|--------|-------|-------------------|
| | North East | North West | Yorkshire & Humber | West Midlands | East Midland | East | South East | South West | London | Wales | |
| Group A | 1.27 | 5.69 | 7.54 | 4.61 | 13.66 | 2.97 | 2.59 | 0.47 | 3.61 | 0.49 | 4.51 |
| B | 0.05 | 0.36 | 0.14 | 2.69 | 1.76 | 0.45 | 0.85 | 0.05 | 13.55 | 0.02 | 2.51 |
| C | 0.03 | 1.15 | 0.71 | 0.69 | 1.45 | 0.58 | 0.52 | 0.57 | 27.71 | 0.16 | 4.45 |
| D | 0.01 | 0.15 | 0.05 | 0.02 | 0.05 | 0.99 | 1.30 | 0.22 | 19.26 | 0.03 | 3.02 |
| E | 1.89 | 1.73 | 2.07 | 1.60 | 1.49 | 0.83 | 1.24 | 0.59 | 8.51 | 1.16 | 2.36 |
| F | 3.77 | 5.84 | 6.85 | 10.55 | 6.62 | 16.92 | 18.78 | 8.07 | 19.00 | 4.36 | 11.38 |
| G | 92.98 | 85.08 | 82.64 | 79.83 | 74.97 | 77.25 | 74.72 | 90.04 | 8.36 | 93.78 | 71.78 |

Table 5. Regional variations in the composition of CELOAC groups.

Regionally all of the minority CELOAC groups except Group A are much more abundant in London. London, as a global city, has exerted a particularly strong pull on economic

migrants. In 2014, a Boston Consulting Group study which surveyed over 200,000 individuals globally found London to be the most desirable city to work in (BCG, 2014). Consequently, in London the White British ethnic group only account for 44.9% of the population, almost half the national average (as expressed in table 4). Consequently much of the city is represented by a mosaic of minority CELOAC groups.

A new classification for London

Given London's distinctive eclectic composition of ethnicities, it is reasonable to analyse it individually as the nation-wide classification may fail to sufficiently discriminate between small areas within the capital city. A London specific CELOAC was also developed therefore, similar to Longley and Singleton's (2014) London specific Output Area Classification.

The England and Wales CELOAC was created with data standardized by the averages for the entire dataset and the k-means clustering did not consider the spatial distribution of OAs. The results were therefore relative to the whole of England and Wales. Using the same set of variables for London OAs only, the data was re-standardised and the clustering was run again to create a new set of 7 groups.

Figure 5 labels the new groups for London and maps their distribution across the capital. The results appear similar to the England and Wales CELOAC upon first glance. The main difference is that two similarly sized South Asia dominated clusters have formed. One also shares higher proportions of South-East Europeans and is concentrated in North East London. The other has a higher proportion of populations of Indian ethnicity and is concentrated between two pockets on both sides of the City. The second notable difference is that the White British group from the London classification has a lower proportion of the White British ethnicity relative to its counterpart from the England and Wales CELOAC. The mixed group from the London classification is more cosmopolitan and is found in areas largely classified as Non-White British in the national classification. This is because groups E and D are more focused in Central and West London as they differ in compositions slightly relative to their national counterparts.

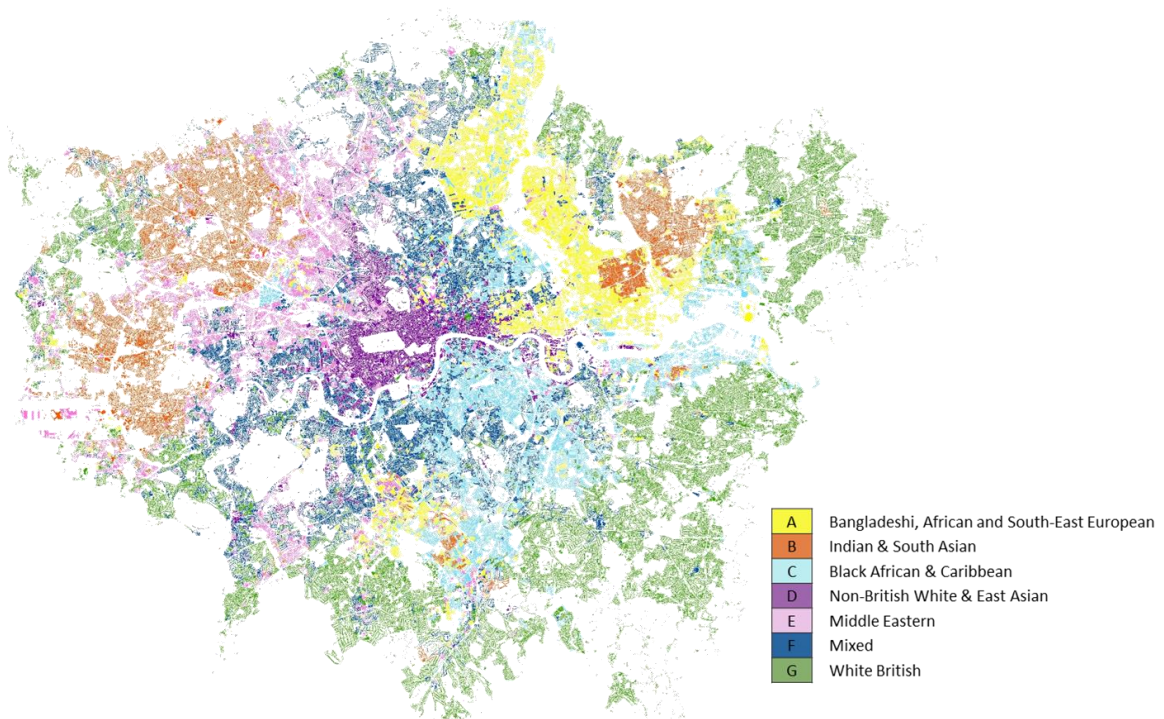


Figure 5. A dasymetric map of the London specific CELOAC. Only areas where buildings are present have been shaded.

Average distances to the cluster centres

The major disadvantage of the K-means clustering method adopted in this project is the potential for cluster distortion, since the algorithm is 'mutually exclusive, collectively exhaustive' and is bound to satisfy the pre-determined value of K' (Debenham 2002: 25). As a result, some OAs might not have been fully optimally clustered as identified by the reclustering of London (figure 5). One way of measuring the uncertainty of the classification is by looking at the average distance of the cases to their cluster centre. The mean distance to the cluster centre for the England and Wales classification is 4.9, which is pretty high considering these values are expressed in Z-scores. Overall, the data is positively skewed, there are relatively few cases which are extremely high above the average. However, it is likely to be due to the nature of ethnic clustering.



Figure 6. The distance of each OA to the cluster centre in England and Wales (left) and London (right). Note: the intervals were rescaled for the London inset.

Figure 6 demonstrates that the distance between the data at each OA and its assigned cluster centre varies across England and Wales. There is a clear urban-rural distinction, urban areas contain much of the instability. Despite the much smaller minority CELOAC groups dominating these areas, individual cultural distinctions mean that many OAs do not fit their clusters as well as many rural OAs fit Group G (White British). There is a notable increase in distances between data and cluster centres and around Thetford, East Anglia. This area has a high proportion of persons of Portuguese original. It is also near to a large RAF base which hosts the largest number of personnel from the United States Air Force in the UK. Within London, there is more uncertainty in areas which became groups A and E in the London classification, as previously these clusters were not well represented in the national version.

Some areas of high instability could also be due to a mutual presence of multiple ethnic groups which are not common in other parts of the country, or due to especially high concentrations of a particular group which may dominate an OA. The classification only considers one main domain of geodemographics, and a relatively volatile one due to the wide range of ethnicities and their tendency to cluster. This tendency is notoriously difficult to measure comprehensively (Massey and Denton, 1988).

Benchmarking CELOAC

Following the development of CELOAC, the classification was benchmarked against grocery store records for selected foods associated with ethnic minorities. Ethnic and cultural identity and heritage can greatly influence consumption, especially food (Kershen, 2002; Jamal, 1998; Hamlett *et al.*, 2008). J Sainsbury's provided the number of sales for six pre-selected grocery products from their stores by customers registered at each OA. The data represents the total number of sales within a 52 week period commencing on 15th May 2011. It was transformed into the proportion of all foods sold and was extracted from the supermarket's internal customer loyalty databases. The foods were chosen due to their distinctive cultural heritage with minority groups. The data was cross tabulated by the whole classification and the results of six foods are shown below as location quotients (table 6).

| Group | Black Eye Beans | Chickpeas | Chinese Leaf | Ghee | Halal | Pickled Cucumbers |
|--------------------------------|-----------------|-----------|--------------|-------|-------|-------------------|
| A: Pakistani and Bangladeshi | 215.9 | 80.99 | 81.49 | 250.5 | 163.2 | 122.0 |
| B: Indian and South Asian Mix | 472.7 | 111.2 | 137.8 | 601 | 711.5 | 312.4 |
| C: Black African & Caribbean | 305.8 | 120.2 | 131.6 | 277.2 | 598.5 | 286.3 |
| D: Non-British White | 218.7 | 230.3 | 229.7 | 216.1 | 413.4 | 356.1 |
| E: Middle Eastern & East Asian | 202.5 | 122.7 | 260.9 | 229.7 | 402.4 | 286.0 |
| F: Mixed | 151.7 | 136.3 | 151.4 | 151.4 | 109.6 | 184.5 |
| G: White British | 50.4 | 87.58 | 78.99 | 44.77 | 19.14 | 49.37 |

Table 6. Location quotients of the rate of world food sales to customers from each of the CELOAC groups.

The data is expressed as an index whereby 100 represents an average representation, values above 100 represent overrepresentation. The results identified substantial variations in consumption across the ethnic groups, notably, the penetration of the selected products is low in Group G (White British). Group B has the highest rates of world food sales, this is most likely because it has the lowest proportion of White British persons. Generally, while the foods may sell particularly well in one minority group, they will often sell better than the national average across all of them reflecting the cosmopolitan composition of their populations.

The results are especially compelling given that migrant groups may be less likely to patronise Sainsbury's stores than the White British population due to its traditional association with the middle class. One must also consider that produce may not be evenly stocked across all of the Sainsbury's store network and food may not be purchased exclusively by its associated cultural group. For example, halal meat is far more likely to be sold in locations with a heavier Muslim presence than in more homogenous white British neighbourhoods.

Conclusion

This study has presented an open-source output area classification of ethnic, cultural and linguistic groups for England and Wales. The research has arisen from the successes of the Output Area Classifications by encompassing only open data from the 2011 Census and utilising a K-means algorithm to cluster OAs. Distinctively, the classification is only composed of variables pertaining to cultural identity, taking full advantage of the highly

granular variable tables made available from the last Census. Unlike geodemographic classifications, it did not consider socio-economics, demographics or other features which describe the population. Despite this difference, the CELOAC revealed a distinctive geography of culturally distinctive neighbourhoods. Whilst rural and many suburban areas largely comprise homogenously White British communities, the inner cities of larger, more globally connected urban areas comprise a more heterogeneous mix of cultural groups. Such groups cluster together and segregate themselves from dissimilar communities to a certain extent, forming spatial mosaics of neighbourhoods in major metropolitan areas, London being a notable example.

There remain opportunities for improvement, the incorporation of additional datasets could be fruitful and could overcome some of the limitations of using data from the 2011 Census. Furthermore, the classification could go into more intricate detail and develop subgroups, and it could expand its scope to include data from Scotland and Northern Ireland. The grocery consumption data provided by a large supermarket chain confirmed that the classification was a good identifier of consumption practices. And it is therefore valid to assume that local ethnic composition is an important part of wider community identity. Furthermore, it proved that such a classification could be useful to planners and analysts from a range of different industries including health, education and retail.

References

- Abrahamson, M. (1995) *Urban Enclaves: Identity and Place in America*. New York: St. Martin's Press.
- Bulmer, M. (1996) 'The ethnic group question in the 1991 Census of population' in Coleman, D. and Salt J. (des.) *Ethnicity in the 1991 Census. vol.1 Demographic characteristics of the ethnic minority populations* HMSO, London
- The Boston Consulting Group (2014) Decoding Global Talent: 200,000 Survey Responses on Global Mobility and Employment Preferences. [Online] www.bcgperspectives.com/content/articles/human_resources_leadership_decoding_global_talent/ (Accessed 23/10/14)
- Cheung, B. Y., Chudek, M and Heine, S.J. (2011) Evidence for a Sensitive Period for Acculturation: Younger Immigrants Report Acculturating at a Faster Rate. *Psychological Science* 22(20) 147-152.
- Debenham, J. (2002) *Understanding Geodemographic Classification: Creating the Building Blocks For An Extension*. Working Paper, School of Geography, University of Leeds. [Online] eprints.whiterose.ac.uk/5014/1/02-1.pdf (Accessed 03/08/14)
- Dustmann, C. Frattini, T. and Theodoropoulos, N. (2011) Ethnicity and Second Generation Immigrants, In Gregg, P. and Wadsworth, J. *The Labour Market in Winter: the state of working Britain 2010*. Oxford: Oxford University Press, Ch 15
- Finney, N. (2013) "How ethnic mix changes and what this means for integration." In van Ham, M., Manley, D., Bailey, N., Simpson, L., MacLennan, D (Eds) *Understanding Dynamic Neighbourhoods*, New York: Springer
- Finney, N. & Simpson, L. (2009) *'Sleepwalking to segregation'? Challenging Myths about Race and Migration*, London: Policy Press
- Flowerdew, R. and Leventhal, B. (1998) Under the microscope, *New Perspectives*, 18: 36-8
- Gale, C.G., Singleton, A. D., Bates, A. G. and Longley, P.A. (2015) Creating the 2011 Area Classification for Output Areas (2011 OAC) Submitted to the Journal of Spatial Information Science.
- Hamlett, J., Bailey A., Alexander, A. and Shaw G. (2008) Ethnicity and Consumption South Asian food shopping patterns in Britain, 1947 – 1975. *Journal of Consumer Culture* 8(1) 91-116.
- Harris R, Sleight P, Webber R. (2005) *Geodemographics: neighbourhood targeting and GIS*. Chichester: John Wiley and Sons.
- Jamal, A. (1998) "Food consumption among ethnic minorities: the case of British-Pakistanis in Bradford, UK", *British Food Journal*, 100(5), 221 - 227

- Johnston, R. J., Poulsen, M.F. & Forrest, J. (2007) The geography of ethnic residential segregation: A comparative study of five countries. *Annals of the Association of American Geographers*, 97, 713-738
- Levinson D. (1998) *Ethnic Groups Worldwide: A Ready Reference Handbook*. New York: Greenwood Press.
- Longley P. A., Cheshire, J, and Mateos, P. (2011) Creating a Regional Geography of Britain through the Spatial Analysis of Surnames, *Geoforum*, 42 (4), 506-516
- Longley, P. A. and Singleton, A. (2014) London Output Area Classification (LOAC): Final Report. GLA Intelligence [Online] londondatastore-upload.s3.amazonaws.com/Vik%3D2011+LOAC+Report.pdf (Accessed 04/11/14)
- Massey, D. S. (1990) The social and economic origins of immigration. *Annals of the American Academy of Political and Social Science* 510, 60-72.
- Massey, D. S. and Denton, N. A. (1988) The dimensions of residential segregation. *Social Forces*, 67(2), 281 - 315
- Mateos P, Singleton A, Longley P (2009) Uncertainty in the analysis of ethnicity classifications: issues of extent and aggregation of ethnic groups. *Journal of Ethnic and Migration Studies* 35(9), 1437-1460
- Mateos, P. (2014a) The international comparability of ethnicity classifications and its consequences for segregation Studies. In: Lloyd C, Shuttleworth I, Wong D (eds) *Social-spatial segregation: concepts, processes and outcomes*. Bristol: Policy Press
- Mateos, P. (2014b) Names, Ethnicity and Populations, *Advances in Spatial Science*, Berlin: Springer-Verlag
- Meyers, L.S., Gamst, G., & Guarino, A.J. (2006). *Applied multivariate research: Design and interpretation*. Thousand Oaks: Sage.
- Milligan, G. W. (1996), Clustering validation: Results and implications for applied analyses, in Arabie, P., Hubert, L. J. and De Soete, G. Eds., *Clustering and Classification*, Singapore, World Scientific.
- Peach, C. (1996) Good segregation, bad segregation. *Planning Perspectives*, 11: 1-20.
- Peach, C. (2006) The mosaic versus the melting pot: Canada and the USA. *Scottish Geographical Journal* 121: 3-27
- Kershen, A. J. ed. (2002) *Food in the Migrant Experience*. Aldershot: Ashgate.
- Portes A, Jensen L.(1987) What's an ethnic enclave? The case for conceptual clarity. *American Sociological Review*, 52:768-771
- Rummel, R. J. (1970) *Applied Factor Analysis*. Evanston: Northwestern University Press
- Simpson, L. (2004) Statistics of Racial Segregation: Measures, Evidence and Policy. *Urban Studies*, 41(3): 661-681.
- Simpson, L. (2013) What makes ethnic group populations grow? *Age structures and immigration. Dynamics of Diversity: Evidence from the 2011 Census*. Centre on Dynamics of Ethnicity. [Online] www.ethnicity.ac.uk/medialibrary/briefings/dynamicsofdiversity/what-makes-ethnic-group-populations-grow-age-structures-and-immigration.pdf (Accessed 02.04.15)
- Simpson, L. and Finney, N. (2009) Spatial Patterns of Internal Migration: Evidence for Ethnic Groups in Britain. *Population, Space and Place*, 15(1): 37-56
- Simpson, L., Jivraj, S. and Warren, J. (2014) The stability of ethnic group and religion in the Censuses of England and Wales 2001-2011. CCSR/CoDE Working Paper 2014 University of Manchester
- Stillwell, J. and Hussain, S. (2010) Internal migration of ethnic groups in Britain, Chapter 5 in Stillwell, J., Finney, N. and Van Ham, M. (eds.) *Understanding Population Trends and Processes Volume 3: Ethnicity and Integration*. Dordrecht: Springer
- Uskul, A. K. and Platt, L (2014) A note on maintenance of ethnic origin diet and healthy eating in Understanding Society. Institute for Social and Economic Research (ISER). Working Paper. [Online] www.iser.essex.ac.uk/publications/working-papers/iser/2014-03 (Accessed 09/11/14)
- Vaughan, L; (1997) *The Urban 'Ghetto': the spatial distribution of ethnic minorities*. In: (Proceedings) First International Space Syntax Symposium. London, UK.
- Vaughan, L. (2007) The spatial foundations of community construction: the future of pluralism in Britain's 'multi-cultural' society. *Global Built Environment Review*, 6 (2) 3-17.

- Vickers, D. and Rees, P. (2006) Introducing the Area Classification of Output Areas. *Population trends*, 125;15-29
- Vickers D and Rees P. (2007). Creating the UK National Statistics 2001 Output Area Classification. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 170(2): 379-403
- Wan, E. and Vanderwerf, M (2009) *A review of the literature on ethnicity, national identity and related missiological studies*. [Online] Available from: www.globalmissiology.org/portugues/docs_pdf/featured/wan_literature_ethnicity_april_2009.pdf (Accessed 08/08/14)